

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: March 14, 2003, 05:42:09 ; Search time 1720 Seconds

(without alignments)  
2429.323 Million cell updates/sec

Title: US-09-698-781-3

Perfect score: 258

Sequence: 1 MKOILPALETTAMTLPVL.....KHQVROSCASCSNSIX 258

Scoring table:

OLIGO  
Xgapop 60.0, Xgapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 1

Total number of hits satisfying chosen parameters: 341630

Minimum DB seq length: 20

Maximum DB seq length: 99

Post-processing: Listing first 100 summaries

Command line parameters:

-MODEL-frame+ .p2n.model -DEV-xih  
-O/cgnt2.1/USPTO.spool/US09698781/runat\_07032003\_083810\_6968/app.query.fasta\_1.455  
-DB-EST -OPTM-fastap -SUFFIX-oli.rst -MINMATCH=0.1 -LOOPEXT=0  
-UNITS-bits -STRAT=1 -END=1 -MATRIX-oligo -TRANS-human40.cdi -LIST=100  
-DOCALIGN=200 -THR.SCOR=quality -THR.MIN=1 -ALIGN=30 -MODE=LOCAL -OUTFMT=pro  
-USBR=US09698781.ecgn\_1.1906.ecgn\_07032003\_083810\_6968 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MAP -LARGEQUERY -NEG.SCOR=0 -WAIT -LONGLOC -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database:

EST:  
1: em\_estda:  
2: em\_esthum:  
3: em\_estlin:  
4: em\_estinu:  
5: em\_estlov:  
6: em\_estlpl:  
7: em\_estro:  
8: em\_hlc:  
9: gb\_est1:  
10: gb\_est2:  
11: gb\_hlc:  
12: gb\_est3:  
13: gb\_est4:  
14: gb\_est5:  
15: em\_estfun:  
16: em\_estom:  
17: gb\_gss:  
18: em\_gss\_hum:  
19: em\_gss\_inv:  
20: em\_gss\_pln:  
21: em\_gss\_vrt:  
22: em\_gss\_fun:  
23: em\_gss\_mam:  
24: em\_gss\_mus:  
25: em\_gss\_other:  
26: em\_gss\_pro:  
27: em\_gss\_rod:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 18    | 7.0         | 55     | 9     | AA661880    |
| 2          | 7     | 2.7         | 32     | 17    | A2961946    |
| 3          | 7     | 2.7         | 81     | 14    | B0761111    |
| 4          | 7     | 2.7         | 92     | 14    | T89904      |
| 5          | 7     | 2.7         | 92     | 17    | BH638251    |
| 6          | 7     | 2.7         | 99     | 9     | AA143118    |
| 7          | 6     | 2.3         | 31     | 9     | A1142775    |
| 8          | 6     | 2.3         | 32     | 17    | TA18D05P    |
| 9          | 6     | 2.3         | 35     | 17    | A2838204    |
| 10         | 6     | 2.3         | 40     | 9     | AA726265    |
| 11         | 6     | 2.3         | 41     | 14    | H94066      |
| 12         | 6     | 2.3         | 41     | 17    | A2813416    |
| 13         | 6     | 2.3         | 46     | 17    | A2412916    |
| 14         | 6     | 2.3         | 48     | 14    | C21082      |
| 15         | 6     | 2.3         | 49     | 9     | A1197659    |
| 16         | 6     | 2.3         | 50     | 9     | AJ281329    |
| 17         | 6     | 2.3         | 50     | 9     | AU102979    |
| 18         | 6     | 2.3         | 51     | 9     | AA616601    |
| 19         | 6     | 2.3         | 51     | 17    | BH218157    |
| 20         | 6     | 2.3         | 52     | 17    | B00049      |
| 21         | 6     | 2.3         | 52     | 17    | DR16K1T     |
| 22         | 6     | 2.3         | 58     | 17    | AL758979    |
| 23         | 6     | 2.3         | 59     | 17    | A2316082    |
| 24         | 6     | 2.3         | 60     | 17    | A2313488    |
| 25         | 6     | 2.3         | 62     | 12    | BG099525    |
| 26         | 6     | 2.3         | 62     | 12    | BF037651    |
| 27         | 6     | 2.3         | 63     | 17    | A2500262    |
| 28         | 6     | 2.3         | 63     | 9     | AA761487    |
| 29         | 6     | 2.3         | 63     | 17    | AF219092    |
| 30         | 6     | 2.3         | 64     | 13    | B1386670    |
| 31         | 6     | 2.3         | 64     | 17    | A2479811    |
| 32         | 6     | 2.3         | 65     | 17    | AL766062    |
| 33         | 6     | 2.3         | 66     | 17    | A2331997    |
| 34         | 6     | 2.3         | 67     | 17    | A2617135    |
| 35         | 6     | 2.3         | 68     | 17    | BH804876    |
| 36         | 6     | 2.3         | 69     | 17    | BH638602    |
| 37         | 6     | 2.3         | 72     | 13    | BM424830    |
| 38         | 6     | 2.3         | 73     | 17    | BH853910    |
| 39         | 6     | 2.3         | 73     | 17    | BH860880    |
| 40         | 6     | 2.3         | 73     | 17    | BH862853    |
| 41         | 6     | 2.3         | 74     | 9     | AA854584    |
| 42         | 6     | 2.3         | 74     | 14    | C21414      |
| 43         | 6     | 2.3         | 75     | 17    | A2506280    |
| 44         | 6     | 2.3         | 75     | 17    | A2847788    |
| 45         | 6     | 2.3         | 75     | 17    | AL762744    |
| 46         | 6     | 2.3         | 76     | 9     | A1540827    |
| 47         | 6     | 2.3         | 76     | 14    | B0816306    |
| 48         | 6     | 2.3         | 77     | 13    | A2480476    |
| 49         | 6     | 2.3         | 78     | 13    | B1517667    |
| 50         | 6     | 2.3         | 78     | 17    | BH847206    |
| 51         | 6     | 2.3         | 79     | 9     | A1789866    |
| 52         | 6     | 2.3         | 79     | 14    | B0994073    |
| 53         | 6     | 2.3         | 79     | 14    | B0094091    |
| 54         | 6     | 2.3         | 79     | 14    | B0094092    |
| 55         | 6     | 2.3         | 80     | 9     | AA889157    |
| 56         | 6     | 2.3         | 80     | 17    | CNS01X08    |
| 57         | 6     | 2.3         | 82     | 9     | AA283715    |
| 58         | 6     | 2.3         | 82     | 10    | AA914126    |
| 59         | 6     | 2.3         | 83     | 14    | B0789779    |
| 60         | 6     | 2.3         | 83     | 17    | BH221159    |
| 61         | 6     | 2.3         | 85     | 9     | AJ283196    |
| 62         | 6     | 2.3         | 85     | 13    | B1758902    |
| 63         | 6     | 2.3         | 85     | 13    | BH802812    |
| 64         | 6     | 2.3         | 85     | 17    | BH802812    |

```

c 65 6 2.3 86 9 AA552526
c 66 6 2.3 86 10 AA262304
c 67 6 2.3 86 14 BQ625486
c 68 6 2.3 86 17 A2466721
c 69 6 2.3 86 17 A2759836
c 70 6 2.3 88 9 AA871948
c 71 6 2.3 88 9 F30846
c 72 6 2.3 89 17 AF219091
c 73 6 2.3 89 17 CDS020RD
c 74 6 2.3 90 9 AA971814
c 75 6 2.3 90 9 AF027915
c 76 6 2.3 90 9 A1470812
c 77 6 2.3 90 17 A2921474
c 78 6 2.3 91 9 AA923763
c 79 6 2.3 91 9 A1968724
c 80 6 2.3 91 13 B1557274
c 81 6 2.3 91 17 A2374702
c 82 6 2.3 93 14 BQ376513
c 83 6 2.3 93 17 AL764388
c 84 6 2.3 94 9 A281343
c 85 6 2.3 94 9 A283110
c 86 6 2.3 94 17 A2941059
c 87 6 2.3 94 17 BH612035
c 88 6 2.3 95 9 A1664297
c 89 6 2.3 95 17 A2603988
c 90 6 2.3 95 9 AA012296
c 91 6 2.3 96 10 AA059816
c 92 6 2.3 96 12 BG111920
c 93 6 2.3 96 17 A2465383
c 94 6 2.3 97 17 BH224140
c 95 6 2.3 97 9 A1583847
c 96 6 2.3 98 9 AB036296
c 97 6 2.3 98 9 A0177039
c 98 6 2.3 98 14 BQ566189
c 99 6 2.3 98 17 AL766046
c 100 6 2.3 98 17 CDS030BK

```

## ALIGNMENTS

```

RESULT 1
AA661880 55 bp mRNA linear EST 16-DEC-1997
LOCUS nt95f06.s1 NCLCGAP_P12 Homo sapiens cDNA clone IMAGE:1206275
DEFINITION similar to TR:E221225 E221225 CYSTEINE-RICH SECRETORY PROTEIN-3
PRECUSOR. (1) ; mRNA sequence.
ACCESSION AA661880
VERSION AA661880.1 GI:2615971
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
JOURNAL NCI-CCAP http://www.nci.nlm.nih.gov/ncicgap.
COMMENT 1 (bases 1 to 55)
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaaps-remail.nih.gov
Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D.,
Rodrigo F. Chuagui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Kitzman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILM, at:
www-bio.illn.gov/dbcp/image/image.html

```

Trace considered overall poor quality  
 Insert length: 364 Std Error: 0.00  
 Seq primer: -40m13 fwd. ET from Amersham  
 High quality sequence stop: 1.

```

FEATURES
Source Location/Qualifiers
1..55
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1206275"
/clone_id="NCLCGAP_P12"
/sex="male"
/tissue_type="metastatic prostate bone lesion"
/lab_host="DH10B"
/note="Vector: PAMP10: mRNA made by metastatic prostate
lesion of the bone. cDNA made by oligo-dT priming.
Non-directionally cloned. Size selected on agarose gel,
average insert size 600 bp. Library made by D. Kitzman,
NIH."
BASE COUNT 15 a 11 c 13 g 16 t
ORIGIN
Alignment Scores:
Pred. No.: 8.14e-09 Length: 55
Score: 18.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.98% Indels: 0
DB: 9 Gaps: 0
US-09-698-781-3 (1-258) x AA661880 (1-55)
QY 188 G1YASnTTPAlAsnARGLuTYrYvalProTYrGIuInGIalalProCysAla 205
Db 2 GGTAAATTGGGCTAATGACATATATGTCCTTATGAAACAGACACCTTGTC 55
RESULT 2
AA2961946 32 bp DNA linear GSS 27-APR-2001
LOCUS clone UUCG2M0230002 R. DNA sequence.
DEFINITION
ACCESSION AA2961946
VERSION AA2961946.1 GI:13833173
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 32)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weis, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0230 row: 0 column: 02
Seq primer: CACACAGAAACACCTATGACC
Class: plasmid ends
High quality sequence stop: 32.
Location/Qualifiers
1..32
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="UUCG2M0230002"
/clone_id="Mouse 10kb plasmid UUCG2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, TI-resistant, F-"

```

/note="vector: pMD22uv. Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD22 (9114/32114/9b/AF19072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 3 a 3 c 13 g 13 t

ORIGIN

Alignment Scores:

|                        |         |               |    |
|------------------------|---------|---------------|----|
| Pred. No.:             | 417     | Length:       | 32 |
| Score:                 | 7.00    | Matches:      | 7  |
| Percent Similarity:    | 100.00% | Conservative: | 0  |
| Best Local Similarity: | 100.00% | Mismatches:   | 0  |
| Query Match:           | 2.71%   | Indels:       | 0  |
| DB:                    | 17      | Gaps:         | 0  |

US-09-698-781-3 (1-258) x A2961946 (1-32)

OY 22 PheleuvaiAlaglyleu 28  
|||||

Db 3 TTTCTGACTGCTTATTG 23

#### RESULT 3

LOCUS B0761111 81 bp mRNA linear EST 26-JUL-2002  
DEFINITION EBRO4\_S0004\_B23\_R root, 3 week, salt-stressed, cv Optic, EBRO4  
Hordeum vulgare cDNA clone EBRO4\_S0004\_B23 5', mRNA sequence.  
ACCESSION B0761111  
VERSION B0761111.1 GI:21969583  
KEYWORDS EST.  
SOURCE Hordeum vulgare.  
ORGANISM Hordeum vulgare.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Hordeum.  
1 (bases 1 to 81)  
Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardie, L., Ramsay, L., Machray, G., Marshall, D. F. M. and Maugh, R.  
Development of Barley Transcriptome Resources  
Unpublished (2001)  
Contact: Maugh R, Marshall DF  
Genome Dynamics/Computational Biology  
Scottish Crop Research Institute  
Invergowrie, Dundee, DD2 5DA, Scotland, UK  
Tel: 00 44 1382 562731  
Fax: 00 44 1382 562426  
Email: estescr1.scri.ac.uk.  
Location/Qualifiers

#### FEATURES

source  
1. 81  
/organism="Hordeum vulgare"  
/cultivar="Optic"  
/db\_xref="taxon:4513"  
/clone="EBRO4\_S0004\_B23"  
/clone\_1lb="root, 3 week, salt-stressed, cv Optic, EBRO4"  
/tissue\_type="root"  
/dev\_stage="3 week"  
/lab\_host="DH10B"  
/note="vector: pSPORT1, site 1: Sal I, site 2: Not I;  
Non-normalised library, directionally cloned into pSPORT1.

Derived from roots of 3 week old salt stressed barley plants. Developed as part of the barley transcriptome resources of BBSRC/SEERAD funded cereal IGF (Investigating Gene Function) project.

BASE COUNT 18 a 25 c 26 g 12 t

ORIGIN

Alignment Scores:

|                        |          |               |    |
|------------------------|----------|---------------|----|
| Pred. No.:             | 1.16e+03 | Length:       | 81 |
| Score:                 | 7.00     | Matches:      | 7  |
| Percent Similarity:    | 100.00%  | Conservative: | 0  |
| Best Local Similarity: | 100.00%  | Mismatches:   | 0  |
| Query Match:           | 2.71%    | Indels:       | 0  |
| DB:                    | 14       | Gaps:         | 0  |

US-09-698-781-3 (1-258) x B0761111 (1-81)

OY 151 HistyThrglnvalValTrp 157  
|||||

Db 59 CACTACACACAGCTGCTGTG 79

#### RESULT 4

LOCUS T89904 92 bp mRNA linear EST 20-MAR-1995  
DEFINITION yella09.r1 Stratagene lung (#937210) Homo sapiens cDNA clone  
IMAGE:117400 5', mRNA sequence.  
ACCESSION T89904  
VERSION T89904.1 GI:718417  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 92)  
Hiller, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B., Chisoe, S., Dietrich, N., Dubuque, T., Favello, R., Gish, W., Hawkins, M., Hultman, M., Kucaba, J., Lacy, M., Le, N., Marquis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Riklin, L., Rohlfing, T., Schlenker, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevas, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Maizra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estescr1.wustl.edu  
Insert size: 1079  
High quality sequence stops: 62 Source: IMAGE Consortium, LIND. This clone is available royalty-free through LIND; contact the IMAGE Consortium ([infodiv1.lind.gov](http://infodiv1.lind.gov)) for further information.  
Insert length: 1079 Std Error: 0.00  
Seq primer: M13RP1  
High quality sequence stop: 62.  
Location/Qualifiers

#### FEATURES

source  
1. 92  
/organism="Homo sapiens"  
/db\_xref="GDB:485689"  
/db\_xref="taxon:9606"  
/clone="IMAGE:117400"  
/clone\_1lb="Stratagene lung (#937210)"  
/sex="male"  
/dev\_stage="72 years"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: Lung; Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dT, normal lung. Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5' GATTCGACGACG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."  
BASE COUNT 34 a 21 c 24 g 13 t

Alignment Scores:

| Pred. No.:             | 1.34e+03 | Length:       | 92 |
|------------------------|----------|---------------|----|
| Score:                 | 7.00     | Matches:      | 7  |
| Percent Similarity:    | 100.00%  | Conservative: | 0  |
| Best Local Similarity: | 100.00%  | Mismatches:   | 0  |
| Query Match:           | 2.71%    | Indels:       | 0  |
| DB:                    | 14       | Gaps:         | 0  |

US-09-698-781-3 (1-258) x T89904 (1-92)

OY 25 AlaglyLeuProSerPhe 31  
 |||||  
 Db 86 GCGGGCTTCCTCTCTTC 66

RESULT 5  
 BH638251  
 LOCUS 1008021D04.1EL\_Y1 1008 - Rescemu Grid 1 Zea mays genomic DNA  
 DEFINITION 92 bp DNA linear GSS 14-FEB-2002  
 ACCESION BH638251  
 VERSION BH638251.1 GI:18662118  
 KEYWORDS GSS.  
 SOURCE Zea mays.  
 ORGANISM Zea mays.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 92)  
 Walbot V  
 Maize genomic sequences found using engineered Rescemu transposon unpublished (2001)  
 CONTACT Walbot V  
 Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 723 8221  
 Email: valbot@stanford.edu  
 Very probable ligation site of ends cut by single endonuclease.  
 Reverse complemented post-ligation sequence from source sequence.  
 Plate: 1008021 row: 9  
 Class: transposon tagged.  
 Location/Qualifiers

FEATURES  
 source  
 1..92  
 /organism="Zea mays"  
 /cultivar="mixed background W23/A18/B73"  
 /db\_xref="taxon:4577"  
 /clone\_lib="1008 - Rescemu Grid 1"  
 /tissue\_type="leaf"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Organ: leaf; Vector: Rescemu (engineered from plasmid backbones); Site\_1: BamHI; Site\_2: BglII; Rescemu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on Rescemu, go to the web site www.zmmb.iastate.edu and follow the links for 'Rescemu'. Grid 1 was grown at Berkeley in 2001. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

BASE COUNT 20 a 21 c 19 g 32 t  
 ORIGIN

Alignment Scores:

| Pred. No.:             | 1.34e+03 | Length:       | 92 |
|------------------------|----------|---------------|----|
| Score:                 | 7.00     | Matches:      | 7  |
| Percent Similarity:    | 100.00%  | Conservative: | 0  |
| Best Local Similarity: | 100.00%  | Mismatches:   | 0  |
| Query Match:           | 2.71%    | Indels:       | 0  |

DB: 17 Gaps: 0

US-09-698-781-3 (1-258) x BH638251 (1-92)

OY 16 LeupheProValLeuPhe 22  
 |||||  
 Db 20 CTATTTCCAGTCTACTATT 40

RESULT 6  
 AA143118  
 LOCUS 99 bp mRNA linear EST 08-NOV-1997  
 DEFINITION z069ell.r1 Stragene pancreas (#937208) Homo sapiens cDNA clone IMAGE:592172.5' similar to gb:X57138\_rnal HISTONE H2B.2 (HUMAN);.  
 mRNA sequence.

ACCESION AA143118  
 VERSION AA143118.1 GI:1712497  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 99)  
 REFERENCE Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Stepien, M., Tan, F., Theisling, B., White, X., Wylie, T., Waterston, R. and Wilson, R.  
 WashU-NCI human EST Project  
 Unpublished (1997)  
 CONTACT: Wilson R  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 Trace considered overall poor quality  
 Insert Length: 622 Std Error: 0.00  
 Seq primer: -28613 rev2 from Amsterdam  
 High quality sequence stop: 1.  
 Location/Qualifiers

FEATURES  
 source  
 1..99  
 /organism="Homo sapiens"  
 /db\_xref="GDB:4623558"  
 /db\_xref="taxon:9606"  
 /clone\_lib="IMAGE:592172"  
 /clone\_lib="Stragene pancreas (#937208)"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 /note="Organ: pancreas; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: 0.010 dt. Pancreatic adenocarcinoma cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACAG 3' -3' adaptor sequence: 5' CTCGACGTTTCTTTTCTTTT 3'."  
 CTCTGACGTTTCTTTTCTTTT 3'.

BASE COUNT 37 a 21 c 27 g 10 t 4 others  
 ORIGIN

Alignment Scores:

| Pred. No.:             | 1.45e+03 | Length:       | 99 |
|------------------------|----------|---------------|----|
| Score:                 | 7.00     | Matches:      | 7  |
| Percent Similarity:    | 100.00%  | Conservative: | 0  |
| Best Local Similarity: | 100.00%  | Mismatches:   | 0  |
| Query Match:           | 2.71%    | Indels:       | 0  |
| DB:                    | 9        | Gaps:         | 0  |

US-09-698-781-3 (1-258) x AA143118 (1-99)

OY 202 AlaprocysAlasercysPro 208  
 |||||  
 Db 66 GCTCGCTGCGCTCTTCGCCG 46

RESULT 7  
 AA142775/c

LOCUS A1142775 31 bp mRNA EST 23-OCT-1998  
 DEFINITION q26a02.s1 Soares\_NHMPU\_S1 Homo sapiens CDNA IMAGE:1687850  
 3' similar to TR:Q05519 Q05519 ARGININE-RICH 54 KD NUCLEAR PROTEIN.  
 ; mRNA sequence.  
 ACCESSION A1142775  
 VERSION A1142775.1 GI:3659134.  
 EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 31)  
 AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
 This clone is available royalty-free through LNC ; contact the  
 IMAGE Consortium ([infoimage.lnl.gov](mailto:infoimage.lnl.gov)) for further information.  
 Trace considered overall poor quality  
 Insert Length: 1957 Std Error: 0.00  
 Seq primer: -40m13 fwd. Ex from Amerisham  
 High quality sequence stop: 1.  
 FEATURES  
 source  
 1. 31  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1687850"  
 /clone\_1b="Soares\_NHMPU\_S1"  
 /tissue-type="Pooled human melanocyte, fetal heart, and  
 pregnant uterus"  
 /lab-host="DH108"  
 /note="Organ: mixed (see below); Vector: pTZ19D-Pac  
 (Pharmacia) with a modified polylinker; Site\_1: Not T;  
 Site\_2: Eco RI; Equal amounts of plasmid DNA from three  
 normalized libraries (melanocyte 2NbM, pregnant uterus  
 NBHPU, and fetal heart NBH19W) were mixed, and ss circles  
 were made in vitro. Following HAP purification, this DNA  
 was used as tracer in a subtractive hybridization  
 reaction. The driver was PCR-amplified cDNAs from pools of  
 5,000 clones made from the same 3 libraries. The pools  
 consisted of 1 M.A.G.E. clones 260232-265223,  
 340488-345479, and 484488-489479."  
 BASE COUNT 7 a 12 c 8 g 4 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 4e+03 Length: 31  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.33% Indels: 0  
 DB: 9 Gaps: 0  
 US-09-698-781-3 (1-258) x A1142775 (1-31)  
 OY 24 ValAlaGlyLeuLeuPro 29  
 Db 25 GTGGCAGGTCTCTGCTCT 8  
 RESULT 8  
 LOCUS TA318D05P 32 bp DNA linear GSS 13-DEC-2000  
 DEFINITION T. brucei sheared genomic DNA clone 318d05, forward sequence.  
 ACCESSION AL492652  
 VERSION AL492652.1 GI:11867479  
 GSS.  
 KEYWORDS Trypanosoma brucei.  
 SOURCE Trypanosoma brucei.  
 ORGANISM Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
 Trypanosoma.

REFERENCE 1 (bases 1 to 32)  
 AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,  
 Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,  
 Melville, S.E., Rajandream, M.A. and Barrell, B.G.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing  
 project, Sanger Centre, The Wellcome Trust Genome Campus,  
 Hinxton, Cambridge CB10 1SA. E-mail: [barrell@sanger.ac.uk](mailto:barrell@sanger.ac.uk) and  
[nilesanger@sanger.ac.uk](mailto:nilesanger@sanger.ac.uk)  
 COMMENT Constructed at the institute for Genomic Research (TIGR),  
 Rockville, MD. Genomic DNA isolated from a cloned population of  
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared  
 to give a tight size distribution (4 kb). The v + i method used for the library construction is  
 described in detail in Smith, H. and Venter, J.C. (Making small  
 insert libraries for whole genome shotgun sequencing projects. In  
 Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.  
 Barrell, Oxford University Press, 1999).  
 Email: [nilesanger@tigr.org](mailto:nilesanger@tigr.org)  
 Details of T. brucei sequencing at the Sanger Centre are available  
 at <http://www.sanger.ac.uk/projects/T-brucei/>.  
 FEATURES  
 source  
 1. 32  
 Location/Qualifiers  
 /organism="Trypanosoma brucei"  
 /strain="TREU927"  
 /db\_xref="taxon:5691"  
 /clone="318d05"  
 BASE COUNT 8 a 6 c 5 g 13 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 4.15e+03 Length: 32  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.33% Indels: 0  
 DB: 17 Gaps: 0  
 US-09-698-781-3 (1-258) x TA318D05P (1-32)  
 OY 228 AsnCysLysSerLeuLys 233  
 Db 23 AACTGTAAGCTTAAG 6  
 RESULT 9  
 LOCUS A2838204/c 35 bp DNA linear GSS 20-FEB-2001  
 DEFINITION 2M0133K17R Mouse 10kb plasmid UGCC1W library Mus musculus genomic  
 clone UGCC2M0133K17 R. DNA sequence.  
 ACCESSION A2838204  
 VERSION A2838204.1 GI:13008112  
 GSS.  
 KEYWORDS house mouse.  
 SOURCE Mus musculus.  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sclurognathu; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 35)  
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Bacon, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Federsen, T., Reilly,  
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
 and Wright, D., Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)  
 Insert Length: 10000 Std Error: 0.00

Plate: 0133 row: K column: 17  
 Seq primer: CACACAGCAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 35.  
 Location/Qualifiers

# FEATURES

source

1..35  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="U08C2M033K17"  
 /clone\_1lb="Mouse 10kb plasmid U08C1M library"  
 /sex="Male"  
 /lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PMD42ny: Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b/AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 11 a 0 c 24 g 0 t

## Alignment Scores:

Pred. No.: 4.58e+03 Length: 35  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.33% Indels: 0  
 DB: 17 Gaps: 0

US-09-698-781-3 (1-258) x A2838204 (1-35)

OY 27 LeuLeuProSerPhePro 32

DB 20 CTCCTCCCTCCCTCCCT 3

RESULT 10

AA726265

LOCUS

DEFINITION

AA726265 40 bp mRNA linear EST 02-JAN-1998  
 v889c01.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone  
 IMAGE:1209312 5' similar to TR:008634 008634 HIGH-SOLPUR KERATIN  
 PROTEIN ; mRNA sequence.

ACCESSION

AA726265 GI:2743972

VERSION

EST.

KEYWORDS

house mouse.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE

1 (bases 1 to 40)

JOURNAL

Marr M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T.,

COMMENT

Geisel S., Kucaba T., Lacey M., Le M., Martin J., Morris M.,

Washington University School of Medicine

Schneidberg R., Stepien M., Tan F., Underwood R., Moore B.,

Washington University School of Medicine

Truesing B., Wylie T., Lennon G., Soares B., Wilson R. and

# FEATURES

source

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:645656  
 putative full length read  
 vector to vector length is 592  
 Seq primer: -28m13 rev1 EF from Amersham  
 High quality sequence stop: 1.  
 Location/Qualifiers

BASE COUNT 3 a 18 c 11 g 8 t

## Alignment Scores:

Pred. No.: 5.31e+03 Length: 40  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.33% Indels: 0  
 DB: 9 Gaps: 0

US-09-698-781-3 (1-258) x AA726265 (1-40)

OY 65 AlaValSerProProAla 70

DB 18 GCAGTGTCCACCGCGCC 35

RESULT 11

H94066

LOCUS

DEFINITION

H94066 41 bp mRNA linear EST 25-NOV-1996  
 yv15h02.s1 Soares fetal liver spleen INFUS Homo sapiens cDNA clone  
 IMAGE:242833 3' similar to gb:Y00493.fna1 HEMOGLOBIN ALPHA CHAIN  
 (HOMAN); mRNA sequence.

ACCESSION

H94066 GI:1101362

VERSION

EST.

KEYWORDS

human.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE

1 (bases 1 to 41)

JOURNAL

Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M., Holman

COMMENT

R., Hultman M., Kucaba T., Le M., Lennon G., Marr M., Parsons J.,

Washington University School of Medicine

R., Williamson A., Wohlmann P. and Wilson R.

Washington University School of Medicine

Unpublished (1995)

Washington University School of Medicine

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Washington University School of Medicine

Tel: 314 286 1800

Washington University School of Medicine

Fax: 314 286 1810

Washington University School of Medicine

Email: est@wustl.edu

Washington University School of Medicine

High quality sequence starts: 1

Washington University School of Medicine

High quality sequence stops: 1

Source: IMAGE Consortium, LNL  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 Trace considered overall poor quality  
 Insert Length: 1147 Std Error: 0.00  
 Seq primer: m13 -40 forward.  
 Location/Qualifiers  
 1. 41

FEATURES  
 source  
 /organism="Homo sapiens"  
 /db\_xref="GDB:3791968"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:242835"  
 /clone.lib="Scores fetal liver spleen INTUS"  
 /sex="male"  
 /dev\_stage="20 week-post conception fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)  
 with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
 1st strand cDNA was primed with a Pac I - oligo(dT) primer  
 [5' AACTGAGAGATTAATTAAGATCTTTTCTTTTCTTTT 3'],  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Pac I and cloned into the Pac I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization. Library  
 constructed by Bento Soares and M.Falima Bonaldo."  
 BASE COUNT  
 ORIGIN  
 9 a 9 c 10 g 11 t 2 others

Alignment Scores:  
 Pred. No.: 5.45e+03 Length: 41  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.33% Indels: 0  
 DB: 14 Gaps: 0

US-09-698-781-3 (1-258) x H94066 (1-41)

OY 230 LysSerLeuLysLeuThr 235  
 Db 2 AAGACCTTGAACTTGACC 19

RESULT 12  
 A2813416 41 bp DNA linear GSS 20-FEB-2001  
 LOCUS 2M0080M10R Mouse 10kb plasmid UGCC1M library Mus musculus genomic  
 DEFINITION  
 A2813416  
 A2813416  
 A2813416.1 GI:12983324  
 GSS.  
 house mouse.  
 SOURCE  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 41)  
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,  
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
 and Wright,D., Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished (2000)  
 CONTACT: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel.: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0080 Row: M Column: 10  
 Seq primer: CACACAGGAACGCTATGACC

Class: plasmid ends  
 High quality sequence stop: 41.  
 Location/Qualifiers  
 1. 41

FEATURES  
 source  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UCC2M0080M10"  
 /clone.lib="Mouse 10kb plasmid UGCC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-gold, T1-resistant, F-"  
 /note="Vector: pMD42ny; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adapted DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pMD42 (g1147321149b1ar129072.1), a copy-number  
 indecible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adapted mouse DNA was annealed to  
 adapted vector DNA, and transformed into  
 chemically-competent E. coli XL10-gold (Stratagene) cells  
 and selected for ampicillin resistance."

BASE COUNT  
 ORIGIN  
 7 a 13 c 8 g 13 t

Alignment Scores:  
 Pred. No.: 5.45e+03 Length: 41  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.33% Indels: 0  
 DB: 17 Gaps: 0

US-09-698-781-3 (1-258) x A2813416 (1-41)

OY 231 SerLeuLysLeuThrLeu 236  
 Db 21 TCCCTCAAGCTGACCTTG 38

RESULT 13  
 A2412916 46 bp DNA linear GSS 03-OCT-2000  
 LOCUS 1M0186J11R Mouse 10kb plasmid UGCC1M library Mus musculus genomic  
 DEFINITION  
 A2412916  
 A2412916  
 A2412916.1 GI:10536929  
 GSS.  
 house mouse.  
 SOURCE  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 46)  
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,  
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
 and Wright,D., Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished (2000)  
 CONTACT: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel.: 801 585 5606

Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0186 row: J column: 11  
 Seq primer: CACACAGCAACACACTATACCC  
 Class: plasmid ends  
 High quality sequence stop: 46.  
 Location/Qualifiers  
 1..46

## FEATURES

source

/organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="U08CJM0186J11"  
 /clone\_1lb="Mouse 10kb plasmid U08CJM library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PMD29v, Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD2 (g114732114[9b]AF19072.1), a copy number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."  
 BASE COUNT 8 a 16 c 9 g 13 t  
 ORIGIN

## Alignment Scores:

Pred. NO.: 6.19e+03 Length: 46  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.33% Indels: 0  
 DB: 17 Gaps: 0

US-09-698-781-3 (1-258) x A2412916 (1-46)

OY 27 LeuLeuProSerPhePro 32

DB 27 CTGCTCCCTCTCCCT 44

## RESULT 14

C21082/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 48)  
 Okubo, K.  
 BodyMap: human gene expression database  
 Contact: Okubo, K.  
 Institute for Molecular and Cellular Biol  
 Osaka University  
 1-3, Yamada-oka, Suita, Osaka Pref. 565, Japan  
 Tel: 06-877-5111(ex.3315)

Email: kouasuke@imb.osaka-u.ac.jp  
 Human Gene Signature, 3'-directed cDNA sequence. We are not  
 submitting the same cDNA sequence redundantly to DBJ since 1993.  
 For the abundance information of clones with this sequence in this  
 library and as well as in other 3'-directed libraries, see  
 http://www.imb.osaka-u.ac.jp/bodymap/. The sequences of the clones  
 represented by this GS sequences is also found there.  
 Location/Qualifiers  
 1..48

## FEATURES

source

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_1lb="Human adult (K.Okubo)"  
 /dev\_stage="adult"  
 /note="Organ: blood; Vector: 1-9t-11; Site: 1; Eco-RI;  
 Monocytes were prepared from blood by ficoll-hypaque,  
 percoll and T cell rosetting purification steps (purity:  
 96 %). mRNA was prepared from activated monocytes from a  
 patient with rheumatoid arthritis. mRNA was reverse  
 transcribed with MuLV. Using Eco-RI linkers cDNA was  
 cloned into 1-9t-11 vector arms. The cDNA library was  
 screened by differential hybridization using radioactively  
 marked ss-cDNA from activated and non-activated  
 monocytes."  
 BASE COUNT 18 a 6 c 9 g 15 t  
 ORIGIN

## Alignment Scores:

Pred. NO.: 6.49e+03 Length: 48  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.33% Indels: 0  
 DB: 14 Gaps: 0

US-09-698-781-3 (1-258) x C21082 (1-48)

OY 123 SerGlnAlaIleGlnSer 128

DB 22 TCACAGCAATTCACAGT 5

## RESULT 15

A1197659/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 49)  
 Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Waterston, R., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 The WashU-HMIT Mouse EST Project  
 Unpublished (1996)  
 Contact: Kara M/Mouse EST Project  
 WashU-HMIT Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box B501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:931788  
 Trace considered overall poor quality  
 Possible reversed clone: similarity on wrong strand



Fri Mar 14 14:00:07 2003

us-09-698-781-3.ol1.rst

Page 9

Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers  
1. .49  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1494184"  
/clone\_11b="Soares-mammary-gland\_NMLMG"  
/sex="female (lactating)"  
/tissue-type="mammary gland"  
/lab-host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; 1st strand cDNA was prepared from mammary  
gland tissue from a lactating female, and was then primed  
with a Not I - O190(OT) primer. Double-stranded cDNA was  
ligated to Eco RI adaptors (Pharmacia), digested with Not  
I and cloned into the Not I and Eco RI sites of the  
modified pT73 vector. Library is normalized. Library  
was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 26 a 13 c 10 g 0 t  
ORIGIN

Alignment Scores:  
Pred. No.: 6.64e+03 Length: 49  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.33% Indels: 0  
DB: 9 Gaps: 0

US-09-698-781-3 (1-258) x A1197659 (1-49)

OY 19 ValLeuNpheuVal 24  
|||||  
Db 42 GTGCTCTGTTCTGCTG 25

RESULT 16  
LOCUS A281329 50 bp mRNA linear EST 30-JUN-2000  
DEFINITION 4A3A-P2A7-F Anopheles gambiae immune competent 4A3A Anopheles  
gambiae cDNA clone 4A3A-P2A7, mRNA sequence.  
ACCESSION A281329  
VERSION A281329.1 GI:6929210  
KEYWORDS EST  
SOURCE African malaria mosquito.  
ORGANISM Anopheles gambiae  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
Anopheles.  
1 (bases 1 to 50)  
Dimopoulos, G., Casavant, T.L., Chang, S., Scheetz, T., Roberts, C.,  
Donohue, M., Schultze, J., Benes, V., Bork, P., Ansoorge, W., Soares, M.B.  
and Kafatos, F.C.  
Anopheles gambiae pilot gene discovery project: identification of  
mosquito innate immunity genes from expressed sequence tags  
generated from immune-competent cell lines  
Proc Natl. Acad. Sci. U.S.A. 97 (12), 6619-6624 (2000)  
Contact: Dimopoulos G  
Fotis C. Kafatos Laboratory  
European Molecular Biology Laboratory  
Meyerhofstrasse 1, 69117 Heidelberg, Germany.  
Location/Qualifiers  
1. .50  
/organism="Anopheles gambiae"  
/strain="4A r/r"  
/db\_xref="taxon:7165"  
/clone="4A3A-P2A7"  
/clone\_11b="Anopheles gambiae immune competent 4A3A"  
/cell\_line="Immune competent 4A3A"  
/lab\_host="E. coli DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: EcoRI; Site\_2: NotI; sequenced from

forward priming site which reads from the 3' end of the  
cDNA. The 4A3A is a directionally cloned and normalized  
cDNA library that was constructed from the 4A3A cell line  
O190-T primed cDNA according to: Bonaldo, Lennon & Soares  
(1996) : Normalization and Subtraction: Two approaches to  
Facilitate Gene Discovery, Genome Research 6, 791-806."

BASE COUNT 19 a 14 c 8 g 8 t 1 others  
ORIGIN

Alignment Scores:  
Pred. No.: 6.79e+03 Length: 50  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.33% Indels: 0  
DB: 9 Gaps: 0

US-09-698-781-3 (1-258) x A281329 (1-50)

OY 232 LeuTyleuThrlleuThr 237  
|||||  
Db 1 CTTAAATTAACCTCACT 18

RESULT 17  
LOCUS A1012979 50 bp mRNA linear EST 30-AUG-2001  
DEFINITION A1012979 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
KAT10408, mRNA sequence.  
ACCESSION A1012979  
VERSION A1012979.1 GI:13552500  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 50)  
Suzuki, Y., Talra, H., Tsunoda, T., Mizushima-Sugano, J., Sese, J., Hata  
H., Ota, T., Isogai, T., Tanaka, T., Morishita, S., Okubo, K., Sakaki  
Y., Nakamura, Y., Suyama, A. and Sugano, S.  
Diverse transcriptional initiation revealed by fine, large-scale  
mapping of mRNA start sites  
EMBO Rep. 2 (5), 388-393 (2001)  
Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yusuzuki@ims.u-tokyo.ac.jp  
Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano  
S. Construction and characterization of a full length-enriched and  
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).  
Location/Qualifiers  
1. .50  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="KAT10408"  
/clone\_11b="Sugano Homo sapiens cDNA library"  
/note="Differential display comparison of untreated and  
dimethylitumate treated 0937 cells"

BASE COUNT 1 a 24 c 14 g 11 t  
ORIGIN

Alignment Scores:  
Pred. No.: 6.79e+03 Length: 50  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.33% Indels: 0  
DB: 9 Gaps: 0

US-09-698-781-3 (1-258) x A1012979 (1-50)

OY 116 SerSerAlaProSerSer 121

Fri Mar 14 14:00:07 2003

us-09-698-781-3.oli.rst

Page 10

```

Db          33      TCCTCCGCCGTCGTCC 50
|||||
RESULT 18
AA616601
LOCUS
DEFINITION
AAG16601
Accession
VERSION
KEYWORDS
SOURCE
ORGANISM
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE
The Mashu-HMI Mouse EST Project
JOURNAL
COMMENT
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel.: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMGE Consortium (info@imge.lnl.gov) for further information.
MG1:576963
Trace considered overall poor quality
Seq primer: -28ml3 rev2 ET from Amerisham
High quality sequence stop: 1.
location/Qualifiers
1..51
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:1026207"
/clone_lib="Barstead mouse proximal colon MPLRB6"
/dev_stage="7 day juvenile"
/lab_host="DH10B"
/note="Vector: pRT73-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTCGACGGCGCCGCTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATCGCATCCTGS], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pRT73 vector.
Library constructed by Bob Barstead."
BASE COUNT
10 a 15 c 9 g 17 t
ORIGIN
Alignment Scores:
Pred. No.: 6.94e+03 Length: 51
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.33% Indels: 0
DB: 9 Gaps: 0
US-09-698-781-3 (1-258) x AAG16601 (1-51)
OY 27 LeuleuprosarPhePro 32
Db 25 TTGTTACGCTCATTCACA 42
RESULT 19
BH218157/c

```

|   |  |               |      |        |                 |
|---|--|---------------|------|--------|-----------------|
| LOCUS                                     | BH218157   | 51 bp         | DNA  | linear | GS5 08-NOV-2001 |
| DEFINITION                                | 1006077D05_2EL_x1 1006 - Rescemu Grid G zea mays genomic, DNA sequence.  |               |      |        |                 |
| ACCESSION                                 | BH218157   |               |      |        |                 |
| VERSION                                   | BH218157.1   | GI:16810590   |      |        |                 |
| KEYWORDS                                  | GS5  |               |      |        |                 |
| SOURCE                                    | zea mays.  |               |      |        |                 |
| ORGANISM                                  | zea mays.  |               |      |        |                 |
| REFERENCE                                 | Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.   |               |      |        |                 |
| AUTHORS                                   | Malbot,V.  |               |      |        |                 |
| TITLE                                     | Maize genomic sequences found using engineered Rescemu transposon  |               |      |        |                 |
| JOURNAL                                   | Unpublished (2001)   |               |      |        |                 |
| COMMENT                                   | Contact: Malbot V<br>Department of Biological Sciences<br>Stanford University<br>855 California Ave, Palo Alto, CA 94304, USA<br>Tel: 650 723 2227<br>Fax: 650 725 8221<br>Email: valbot@stanford.edu<br>Possible ligation site of ends cut by 2 different endonucleases.<br>Reverse complemented post-ligation sequence from source sequence.<br>Plate: 1006077 row: 27<br>Class: transposon-tagged.<br>Location/Qualifiers<br>1..51<br>/organism="zea mays"<br>/cultivar="mixed background W23/A188/B73"<br>/db_xref="taxon:4577"<br>/clone_lib="1006 - Rescemu Grid G"<br>/tissue_type="leaf"<br>/dev_stage="adult"<br>/lab_host="DH10B"<br>/note="Organ: leaf; Vector: Rescemu (engineered from p Bluescript backbone); Site.1: BamHI; Site.2: BglII; Rescemu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on Rescemu, go to the web site 'www.zmdb.iastate.edu' and follow the links for 'Rescemu'. Grid G was grown at Stanford in 2000. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin." |               |      |        |                 |
| BASE COUNT                                | g a  | 14 c          | 20 g | 8 t    |                 |
| ORIGIN                                    |  |               |      |        |                 |
| Alignment Scores:                         |  |               |      |        |                 |
| Pred. No.:                                | 6.94e+03   | Length:       | 51   |        |                 |
| Score:                                    | 6.00   | Matches:      | 6    |        |                 |
| Percent Similarity:                       | 100.00%  | Conservative: | 0    |        |                 |
| Best Local Similarity:                    | 100.00%  | Mismatches:   | 0    |        |                 |
| Query Match:                              | 2.33%  | Indels:       | 0    |        |                 |
| DB:                                       | 17   | Gaps:         | 0    |        |                 |
| US-09-698-781-3 (1-258) x BH218157 (1-51) |  |               |      |        |                 |
| QY 117                                    | SerialProserSerrTP 122   |               |      |        |                 |
| Db 37                                     | TCTGCACCATCTCTCTGG 20  |               |      |        |                 |
| RESULT 20                                 |  |               |      |        |                 |
| LOCUS                                     | B00049   | 52 bp         | DNA  | linear | GS5 13-JUL-1996 |
| DEFINITION                                | CSRL-100h6-u CSRL flow sorted Chromosome 11 specific cosmid Homo sapiens genomic clone CSRL-100h6, DNA sequence.   |               |      |        |                 |
| ACCESSION                                 | B00049   |               |      |        |                 |
| VERSION                                   | B00049.1   | GI:1409327    |      |        |                 |
| KEYWORDS                                  | GS5  |               |      |        |                 |
| SOURCE                                    | human.   |               |      |        |                 |

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 52)  
Evans, G.A., Burbee, D., Davies, C., Habner, L., Oliver, T., Gilbert, M.,  
Jones, D., Ward, T., Gillian, E., Schagemann, J., Probst, S., Harris,  
J., Deford, J., McFarland, J., Burzinski, K., Khan, M., Kupfer, K. and  
Gartner, H.R.  
Genomic Sequence Sampled Map of Chromosome 11  
Unpublished (1996)  
Contact: Evans GA, Shane Probst  
McBermott Center for Human Growth and Development  
University of Texas Southwestern Medical Center At Dallas  
5323 Harry Hines Blvd, Dallas TX 75235-8591  
Tel: 214-648-1600  
Fax: 214-648-1666  
Email: gevas@utsw.swned.edu, shane@mcdermott.swned.edu  
Seq primer: 17  
Class: cosmid ends  
High quality sequence stop: 52.  
Location/Qualifiers  
1. 52  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CSRU-100h6"  
/clone\_11b="CSRU flow sorted Chromosome 11 specific  
cosmid"  
/sex="female"  
/cell\_type="chimeric hamster somatic cell hybrid"  
/note="Vector: SCOS-1; Human Chromosome 11 specific cosmid  
library prepared from flow sorted human Chromosome 11  
derived from Chinese Hamster Ovary (CHO) monochromosomal  
somatic cell hybrid, 31"

BASE COUNT 10 a 10 g 25 t 1 others  
ORIGIN

Alignment Scores:  
Pred. No.: 7.09e+03 Length: 52  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.33% Indels: 0  
DB: 17 Gaps: 0

US-09-698-781-3 (1-258) x B00049 (1-52)  
OY 26 glyLeuLeuProSerPhe 31  
Db 23 GGCCTTGGCCGCTCTTT 40

RESULT 21  
DR16KIT/c 57 bp DNA linear GSS 06-JUN-2002  
LOCUS Danio rerio genomic clone Dkey-16k1, genomic survey sequence.  
DEFINITION AL733018  
VERSION AL733018.1 GI:21341182  
KEYWORDS GSS;  
SOURCE zebrafish.  
ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 57)  
Humphray, S.J., Huckle, E. and Hunt, S.E.  
Submitted (06-JUN-2002) The Sanger Institute, Wellcome Trust Genome  
Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:  
humquays@sanger.ac.uk Unpublished  
this sequence was generated from the T7 end of BAC 16k1. 16k1 is  
part of the Daniokey Pilot BAC Library created by R. Plasterk and  
N.V. Keygene.  
Further details: [http://www.sanger.ac.uk/Projects/D\\_rerio/](http://www.sanger.ac.uk/Projects/D_rerio/).  
Location/Qualifiers

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

FEATURES  
source

source 1. 57  
/organism="Danio rerio"  
/db\_xref="taxon:7955"  
/clone="Dkey-16k1"  
/tissue\_type="testis"  
/note="vector pindigBAC-536"

BASE COUNT 11 a 17 c 12 g 17 t  
ORIGIN

Alignment Scores:  
Pred. No.: 7.85e+03 Length: 57  
Score: 6.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.33% Indels: 0  
DB: 17 Gaps: 0

US-09-698-781-3 (1-258) x DR16KIT (1-57)  
OY 24 ValAGlyLeuLeuPro 29  
Db 44 GTGGCTGGACTGCTGCCA 27

RESULT 22  
AL758979/c 58 bp DNA linear GSS 18-JUN-2002  
LOCUS Arabidopsis thaliana T-DNA flanking sequence GK-164E11-013275.  
DEFINITION genomic survey sequence.  
ACCESSION AL758979  
VERSION AL758979.1 GI:21497327  
KEYWORDS GSS;  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
1  
Strizhov, N., Li, Y., Rosso, M., Viehoveer, P., Dekker, K., Saedler, H.  
and Weisshaar, B.  
A pipeline for automated high-throughput generation of FSTs  
(flanking sequence tags) from Arabidopsis thaliana T-DNA  
transformed lines  
unpublished  
2  
Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weisshaar, B.  
A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)  
for flanking sequence tag based reverse genetics  
unpublished  
3 (bases 1 to 58)  
Strizhov, N., Rosso, M., Li, Y. and Weisshaar, B.  
Direct Submission  
Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer  
Zuechtungsforshung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany  
This sequence is recovered from the left border of the T-DNA. It  
indicates an insertion close to or within gene At4g30820. The  
sequences are generated at the MPI for Plant Breeding Research in  
the context of the GABI-Kat project. GABI-Kat is part of the German  
Plant Genomics program designated 'GABI'. Information on line  
availability can be found at:  
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.  
Location/Qualifiers

FEATURES  
source  
1. 58  
/organism="Arabidopsis thaliana"  
/strain="Columbia 0"  
/db\_xref="taxon:3702"  
/clone="GK-164E11-013275"  
/clone\_11b="Arabidopsis thaliana T-DNA insertion lines"  
/note="PCR was performed on DNA from Arabidopsis thaliana  
plants (T1) which were transformed with the T-DNA from  
vector pAC161. The lines contain one or more T-DNA  
insertions. The DNA fragment(s) resulting from the PCR  
were directly sequenced to determine the genomic sequence  
flanking the insertion. Sequences displaying significant

similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed.

BASE COUNT 12 a 10 c 18 g 18 t  
 ALIGNMENT SCORES:  
 Pred. No.: 8e+03 Length: 58  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.33% Indels: 0  
 DB: 17 Gaps: 0

US-09-698-781-3 (1-258) x AL758979 (1-58)

OY 37 LysAspProAlaPheThr 42  
 DB 36 AAGGATCCGCTTTCACA 19

RESULT 23  
 A2316082 59 bp DNA linear GSS 29-SEP-2000  
 LOCUS 1M0033P01R Mouse 10kb plasmid UGCCIM library Mus musculus genomic  
 DEFINITION clone UGCCIM0033P01 R, DNA sequence.

ACCESSION A2316082  
 VERSION A2316082.1 GI:10363551  
 KEYWORDS GSS.

SOURCE house mouse.  
 ORGANISM Mus musculus.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 59)  
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausen, A., and Wright, D., Weis, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00  
 Plate: 0033 row: P column: 01  
 Seq primer: CACACAGAAACACCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 59.  
 Location/Qualifiers

FEATURES  
 source

1..59  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UGCCIM0033P01"  
 /clone\_1kb="Mouse 10kb plasmid UGCCIM library"  
 /sex="Male"  
 /lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g114732114/gb1AP129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 19 a 14 c 13 g 13 t  
 ALIGNMENT SCORES:  
 Pred. No.: 8.15e+03 Length: 59  
 Score: 6.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.33% Indels: 0  
 DB: 17 Gaps: 0

US-09-698-781-3 (1-258) x A2316082 (1-59)

OY 204 CysAlaSerCysProAsp 209  
 DB 50 TGTGCTCTGTGCTGCGAT 33

RESULT 24  
 A2313488 60 bp DNA linear GSS 29-SEP-2000  
 LOCUS 1M0029E08R Mouse 10kb plasmid UGCCIM library Mus musculus genomic  
 DEFINITION clone UGCCIM0029E08 R, DNA sequence.

ACCESSION A2313488  
 VERSION A2313488.1 GI:10358436  
 KEYWORDS GSS.

SOURCE house mouse.  
 ORGANISM Mus musculus.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 60)  
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausen, A., and Wright, D., Weis, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00  
 Plate: 0029 row: E column: 08  
 Seq primer: CACACAGAAACACCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 60.  
 Location/Qualifiers

FEATURES  
 source

1..60  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UGCCIM0029E08"  
 /clone\_1kb="Mouse 10kb plasmid UGCCIM library"  
 /sex="Male"  
 /lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMDA2 (g114732114181a1f129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 23 a 0 c 37 g 0 t  
ORIGIN

## Alignment Scores:

|                        |         |               |    |
|------------------------|---------|---------------|----|
| Pred. No.:             | 8.3e+03 | Length:       | 60 |
| Score:                 | 6.00    | Matches:      | 6  |
| Percent Similarity:    | 100.00% | Conservative: | 0  |
| Best Local Similarity: | 100.00% | Mismatches:   | 0  |
| Query Match:           | 2.33%   | Indels:       | 0  |
| DB:                    | 17      | Gaps:         | 0  |

US-09-698-781-3 (1-258) x AZ13488 (1-60)

OY 27 LeuLeuprospherPro 32  
|||||  
DB 55 CTCCTCCCTCCTCCCT 38

## RESULT 25

BC099525 62 bp mRNA linear EST 29-JAN-2001  
LOCUS na535309.x1 NCI-CGAP-Co27 Homo sapiens cDNA clone IMAGE:4205200 3',  
DEFINITION mRNA sequence.

ACCESSION BC099525  
VERSION BC099525.1 GI:12594842  
KEYWORDS EST.

## SOURCE

human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 62)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaaps-remail.nih.gov

CDNA Library Preparation: David B. Kitzman, Ph.D.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LNL  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL, send email to:  
info@image.llnl.gov  
Seq primer: -40up from GIBCO.

FEATURES  
source location/Qualifiers

1..62  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="IMAGE:4205200"  
/clone\_lib="NCI-CGAP-Co27"  
/tissue\_type="adenocarcinoma (mucinous component)"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pAMP1; mRNA made from colonic  
adenocarcinoma, cDNA made by oligo-dT priming.  
Directionally cloned into upc sites. Size-selected on  
agarose gel, average insert size 300 bp. Primary library.  
CDNA Library Preparation: David B. Kitzman, Ph.D.  
Reference: Kitzman et al. (1996) Cancer Research  
56:5360-5385."

BASE COUNT 17 a 16 c 12 g 17 t  
ORIGIN

Alignment Scores:  
Pred. No.: 8.61e+03  
Score: 6.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 2.33%  
DB: 12  
Gaps: 0

US-09-698-781-3 (1-258) x BC099525 (1-62)

OY 252 AsnCysSerAsnSerIle 257  
|||||  
DB 10 AATGCTCGAATTCGATC 27

## RESULT 26

BF037661 62 bp mRNA linear EST 20-OCT-2000  
LOCUS 601462102P1 NIH\_MGC\_66 Homo sapiens cDNA clone IMAGE:3865296 5',  
DEFINITION mRNA sequence.

ACCESSION BF037661  
VERSION BF037661.1 GI:10745972  
KEYWORDS EST.

## SOURCE

human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 62)  
AUTHORS NIH-MGC http://mgs.ncbi.nlm.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgaaps-remail.nih.gov

Tissue Procurement: DCTP/DRP  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov

Plate: LLM6908 row: d column: 01  
High quality sequence start: 15  
High quality sequence stop: 59.

## FEATURES

location/Qualifiers

1..62  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="IMAGE:3865296"  
/clone\_lib="NIH\_MGC\_66"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: ovary; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.8 kb. Library constructed by Life  
Technologies."

BASE COUNT 14 a 18 c 19 g 11 t  
ORIGIN

## Alignment Scores:

|                        |          |               |    |
|------------------------|----------|---------------|----|
| Pred. No.:             | 8.61e+03 | Length:       | 62 |
| Score:                 | 6.00     | Matches:      | 6  |
| Percent Similarity:    | 100.00%  | Conservative: | 0  |
| Best Local Similarity: | 100.00%  | Mismatches:   | 0  |
| Query Match:           | 2.33%    | Indels:       | 0  |
| DB:                    | 12       | Gaps:         | 0  |

US-09-698-781-3 (1-258) x BF037661 (1-62)

OY 19 VallLeuPhenylVal 24  
|||||  
DB 46 GTATTATGTTCTAGTA 29

## RESULT 27



```

RESULT 29
AF219092/c 63 bp DNA linear GSS 17-APR-2000
LOCUS AF219092
DEFINITION AF219092 Human Homo sapiens genomic clone V(Uc9), DNA sequence.
ACCESSION AF219092
VERSION AF219092.1 GI:7581538
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 63)
AUTHORS Hamshire,M., Cross,S., Daniels,M., Lennon,G. and Brook,J.D.
TITLE A transcript map of a 10-Mb region of chromosome 19: A source of
genes for human disorders, including candidates for genes involved
in asthma, heart defects, and eye disorders
JOURNAL Genomics 63 (3), 425-429 (2000)
MEDLINE 20171383
COMMENT Contact: Hamshire M
Institute of Genetics
University of Nottingham
Queen's Medical Center, Nottingham, NG7 2LT, United Kingdom
Class: exon-trapped.
FEATURES
source
1. 63
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="V(Uc9)"
/note="Vector: pMOS Blue"
BASE COUNT 20 a 8 c 25 g 10 t
ORIGIN
Alignment Scores:
Pred. No.: 8.76e+03 Length: 63
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.33% Indels: 0
DB: 17 Gaps: 0
US-09-698-781-3 (1-258) x AF219092 (1-63)
Oy 116 SerSerAlaProSerSer 121
Db 46 TCTCTGCCCCATCATCT 29
RESULT 30
BI388670 64 bp mRNA linear EST 14-DEC-2001
LOCUS BI388670
DEFINITION EST-CD34NN-038 cDNA library from human CD34+ stem/progenitor cells
ACCESSION BI388670
VERSION BI388670.1 GI:17737253
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 64)
AUTHORS Zhou,G., Chen,J., Lee,S., Terry,C., Rowley,J.D. and Wang,S.M.
TITLE The pattern of gene expression in human hematopoietic CD34+
stem/progenitor cells
JOURNAL Unpublished (2001)
COMMENT Contact: Wang SM
Hem/Onc
University of Chicago Medical Center
5841 S. Maryland Ave., MC2115, Chicago, IL 60637, USA
Tel: 773-702-6788
Fax: 773-702-3002
Email: swang1@midway.uchicago.edu
This EST fragment was amplified from cDNA library of human CD34+
stem/progenitor cells with GLDI technique (Generation of Longer

```

```

cDNA fragments from SAGE tags for Gene Identification, Proc. Natl.
Acad. Sci. USA 97, 349, 2000), which starts from the 3' end till
the first CATG site of the targeted cDNA sequence.
Seq primer: M13 Forward.
FEATURES
source
1. 64
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="cDNA library from human CD34+ stem/progenitor
cells"
/tissue_type="Bone marrow"
/cell_type="CD34+ stem/progenitor cells"
/note="3-ESTs converted from the SAGE tag sequences using
GLDI method"
BASE COUNT 8 a 21 c 16 g 19 t
ORIGIN
Alignment Scores:
Pred. No.: 8.92e+03 Length: 64
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.33% Indels: 0
DB: 13 Gaps: 0
US-09-698-781-3 (1-258) x BI388670 (1-64)
Oy 26 GlyLeuLeuProSerPhe 31
Db 26 GGGCTCTCTCCACGCTT 43
Search completed: March 14, 2003, 07:14:15
Job time : 1730 secs

```